Subsetting and Formatting Landsat 7 L0R ETM+ Data Products

Michael R. Reid

Computer Sciences Corporation, Lanham, Maryland 20706 USA

ABSTRACT

The Landsat 7 Processing System (LPS) processes Landsat 7 Enhanced Thematic Mapper Plus (ETM+) instrument data into large, contiguous segments called "subintervals." The LPS-processed subinterval products must be subsetted and reformatted before the Level 1 processing systems can ingest them. The initial full subintervals produced by the LPS are stored mainly in HDF Earth Observing System (HDF-EOS) format, an extension to the Hierarchical Data Format (HDF). The final L0R products are stored in native HDF. The HDF and HDF-EOS application programming interfaces (APIs) can be used for extensive data subsetting and data reorganization. How the HDF and HDF-EOS APIs may be used to efficiently subset, format, and organize Landsat 7 L0R data to create various configurations of L0R products is discussed.

Keywords: Subsetting, Landsat 7, DAAC, HDF, HDF-EOS

1. INTRODUCTION

The Landsat 7 spacecraft transmits CCSDS-encoded ETM+ science data to the receiving ground stations during each fourteen-minute contact. The Landsat Processing System (LPS) decodes the telemetry, divides the contact into subintervals, performs Level Zero (L0) processing, organizes the subintervals into scenes based on the Worldwide Reference System (WRS), and stages the subintervals for archiving at the USGS' EROS[*] Data Center (EDC) Distributed Active Archive Center (DAAC). The LPS receives the wideband instrument data on two channels and renders them into two corresponding formats, "format 1" or "format 2." A separate LPS "string" processes each format. The format 1 data consist of ETM+ imagery from five spectral bands (bands 1-5) and one thermal band (band 6 low-gain). The format 2 data consist of ETM+ imagery from one spectral band (band 7), one thermal band (band 6 high-gain), and one high-resolution panchromatic band (band 8). Both formats also contain Mirror Scan Correction Data (MSCD), which are mainly data quality information; Payload Correction Data (PCD), which consist of spacecraft attitude, ephemeris, and miscellaneous engineering data; and Internal Calibrator (IC) data, which consist of internal lamp and shutter data for the spectral and panchromatic bands and blackbody radiance and shutter information for the thermal band.

A Landsat scene is the set of data, which covers the area of the Earth's surface bounded within a WRS scene. The WRS is a global grid system of intersecting paths and rows used by the Landsat projects, which divide the Earth's surface into a set of quadrants. A full WRS scene covers an area on the ground 185 km wide and 170 km long and the spacecraft images a full WRS scene with 375 scans of the ETM+ instrument's mirror in about twenty-seven seconds (this includes twenty scans of overlap with each of the adjacent scenes). A partial subinterval, also called a "floating scene," is a contiguous segment of data within a subinterval, which is not bounded by WRS rows. It could contain all of the data in the full subinterval; however, the Level 1 (L1) processing systems are not designed to handle segments of data larger than three WRS scenes, so partial subintervals will not normally be larger than that.

Prior to ingest by the Level 1 processing systems, the desired data set must be subsetted out of the LPS subinterval product and reformatted in accordance with the Landsat 7 Level Zero-Reformatted (L0R) product specification.[1] To avoid confusion, in this paper the author shall refer to the initial products generated by the LPS,[2] as simply "the LPS products" and the subsetted L0R distribution products as "the L0R products."

Correspondence: Email: mreid@csc.com

[*] Earth Resources Observation System.

2. THE PRODUCTION OF LANDSAT 7 DATA PRODUCTS

Landsat 7 data products are distributed in the Hierarchical Data Format (HDF), version 4[†]. HDF is a platform-independent file format and function library for storing scientific data developed by the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign.[3] NASA's Earth Observing System Data and Information System (EOSDIS) project has developed an extension to HDF called HDF-Earth Observing System (HDF-EOS).[4] The HDF-EOS structures are built on native HDF structures and can coexist with them in the same HDF file. The Earth Observing System (EOS) project requires that Landsat 7 and all other missions, which use the EOSDIS Core System's (ECS) data archiving and distribution services, produce their initial subsettable science data products in HDF-EOS format. The HDF-EOS application program interface (API) provides functions for storing and subsetting geolocated science data.

The LPS writes out its ETM+ image data (i.e., the Band data), IC data, and MSCD in HDF-EOS format. It writes out the PCD and the multi-band Browse files in native HDF. It writes its metadata as simple ASCII text. Once the LPS has produced its products for a given subinterval, they are archived at the EDC DAAC. When a customer orders a Landsat 7 data product, a system at the EDC DAAC locates the appropriate LPS subinterval, uses the HDF-EOS API to subset out the ordered data product, and writes it to native HDF files in accordance with the L0R specification. The product may then be passed on to the Level 1 processing systems, which apply various corrections to the image data.

3. HDF AND THE HDF-EOS EXTENSIONS

HDF is both a file format and a software package. The software package is a library of functions, which implements a set of APIs. It defines numerous structures for storing science data and provides a large and powerful set of functions for reading, writing and manipulating those data. Data can be written to numerous individual data objects within the same HDF file or they can be stored in a set of separate, but associated external element files. The HDF-EOS extends the native HDF by adding some additional data structures and APIs, which are designed specifically for EOS data. Of interest to Landsat 7 users are the native HDF Scientific Data Set (SDS), Vdata, Vgroup, and the 24-Bit Raster Image (RIS24) data models and the HDF-EOS Swath, Point, and Geolocation Field data models. Both the HDF and the HDF-EOS support more data models than these; however, only those structures used in the Landsat 7 products are discussed here. See Fortner[5] for a much more complete overview of HDF.

3.1. Native HDF Structures Used in the Landsat 7 Products

The HDF API provides a rich set of functions for creating, writing, reading, and manipulating its data objects. Internally, HDF keeps extensive referencing and indexing information about the objects it stores and updates it whenever an object is created or modified. It also has an optional feature, which allows the contents of HDF objects to be stored externally from the main HDF file. The HDF API provides the *HXsetcreatedir* and *SDsetexternalfile* functions to do this. Such file sets consist of a main HDF Directory file, which contains all of the HDF referencing and indexing information, and a set of external element files. The external element files are separate from, but not independent of, the HDF Directory file. Any application, which reads or writes HDF objects accesses them through the Directory file,[‡] so that the data appear to be contained in a single, logical file. The HDF access functions will find the proper external element files and retrieve the contents of those data objects. The L0R products employ this scheme.

3.1.1. The Scientific Data Set (SDS)

A SDS is basically a multidimensional array of numeric data with some associated internal labels and indexing information. When creating a SDS, the calling program specifies its rank and size and can optionally assign one of the dimensions to be of unlimited size. The HDF API provides functions for reading SDS objects in their entireties or by slices and for subsampling and annotating them. The SDS is the HDF structure of choice for writing instrument data and is the one used by the L0R systems for storing the ETM+ image and IC data.

[†] NCSA has now released HDF 5, which uses a different data model than HDF 4.

[‡] In order for most applications to read externally stored data, the external element files must be in the same file system directory as the HDF Directory file or the $HDFEXTDIR environment variable must be set. Often, the application must be run from within the file system directory where the L0R product is located.

3.1.2. The Vdata

A Vdata is a set of ordered fields and records. It resembles a conventional relational database table. The fields can be of differing data types, including non-numeric types and arrays. The PCD, MSCD, scan line offsets, and metadatas are stored in Vdatas in the L0R product.

3.1.3. The Vgroup

A Vgroup is analogous to a directory or folder in a computer's file system. It does not store science data itself, but rather serves as a container in which other HDF data objects, such as SDSs or Vdatas, can be stored. It is used for grouping other data objects within an HDF file. The Vgroup is useful for organizing sets of HDF objects into a hierarchical structure.

3.1.4. The Raster Image Set (RIS24)

HDF provides several data structures for storing color images, one of which is the 24-bit RIS24. The RIS24 is a conventional, true color image model. It forms each 24-bit color pixel in the raster image from three 8-bit monochrome pixels, which will be represented as red, green, and blue. Optionally, the image may be JPEG-compressed. The LPS multi-band Browse files contain JPEG-compressed RIS24 objects.

3.2. HDF-EOS Structures Used in the LPS Products

The HDF-EOS data model defines higher-level, but more specialized, data objects than those in native HDF. The HDF-EOS objects are based on the lower-level native HDF structures and retain a degree of compatibility with them. HDF-EOS files *are* HDF files, but programs designed to read only native HDF will likely not be able to interpret the more complex HDF-EOS objects in a meaningful way. In the case of Landsat 7, only the initial data products that the LPS produces contain HDF-EOS structures. The end users of Landsat 7 data will not encounter these objects in the subsetted L0R products obtained from the EDC DAAC.

3.2.1. The Swath

The Swath data model is based on the SDS. It is a higher-level array structure, which extends the SDS by adding some indexing information so that the data can be subsetted and associated with geolocation data. As the name implies, the Swath data model is designed to reflect the swath traced on the surface of the Earth by an orbiting, Earth observing satellite. Typical of Earth observing satellites, the Landsat 7 spacecraft has a mirror which rapidly scans back and forth perpendicular to the direction the vehicle is moving and reflects the ground image onto the instrument's detectors. The Swath has two dimensions, the along-track dimension, which corresponds to the spacecraft's trajectory and, perpendicular to it, the cross track dimension, which corresponds to the direction the instrument scans. HDF-EOS Swaths can have a third, vertical dimension as well, but the LPS does not use this.

3.2.2. The Point

The Point data model is based on the Vdata, but is a higher-level and more complex table-type structure. Like the Swath, it is subsettable and can be demarcated with geolocation data. It is intended to contain data associated with geolocated points on the surface of the Earth; however, it is flexible enough to store other types of tabular data as well.

3.3.3. The Geolocation Fields

Geolocation data locate Swath or Point data either spatially or temporally. In an HDF-EOS file, each Swath or Point can be accompanied by a Geolocation Field, which contains the geolocation data. Typically, the geolocation data will be an array of ground coordinates or times. In the case of Swaths, the Geolocation Field will be associated with each scan or set of consecutive scans. In the case of a Point, the Geolocation Field will be associated with each record or set of consecutive records. The HDF-EOS also defines a special Geolocation Field called an Index, which is usable with Swaths. In the case of the LPS products, it maps sections of Swaths to scenes, including the irregularly sized partial scenes, which occur at the beginnings and ends of subintervals. The Index allows an application to subset a full or partial WRS scene out of a Swath simply by specifying the geographic coordinates of the scene's corners. The ability to geolocate acquired data makes the HDF-EOS especially useful for earth sciences applications.

4. THE LANDSAT 7 DATA PRODUCTS

The Landsat 7 data products appear in two forms as they are processed through to the L1 state. As described earlier, the LPS produces its data sets on a subinterval basis and stores the image, calibration, and mirror scan data in HDF-EOS format. The EDC DAAC subsets the LPS data and produces an L0R data set, which it stores in native HDF. The L0R product may then be forwarded to the Level 1 systems for further processing. The products generated by the Level 1 systems mirror the L0R products in terms of format.

4.1. The Landsat 7 Processing System (LPS) Products

4.1.1. The Band Files

The Band files, also called "Swath files," contain the image data collected by the ETM+ instrument and are the primary files in the LPS product. The LPS writes them out in HDF-EOS format and use HDF-EOS data structures for storing the imagery and ancillary data. Each file contains the data for a single band.[**] There are six Band files (bands 1-5, 6L) in the format 1 product and three (bands 6H, 7, 8) in the format 2. The primary data structure in each Band file is a named HDF-EOS Swath, which contains the ETM+ image data for the given band. The Band files also contain geolocation information and ancillary data associated with the Swath. The LPS writes the image data to a Swath one scan at a time as it processes the raw data downlinked from the satellite. Each scan contains 8, 16, or 32 data lines depending on the number of ETM+ detectors allocated to the given band. The image is stored in the Swath as a set of consecutive pixel lines, which do not retain the sense of individual scans; however, knowledge of the scan associations is important, so the LPS preserves it in ancillary data.

Each 30m-resolution data line is 6,600 pixels (bytes) in length, including margins. The length of the Swath depends upon the duration of the subinterval. A 30m-resolution band (i.e., bands 1-5, and 7) can produce a maximum of 187,600 data lines (35 scenes worth) in a subinterval. Therefore, a 30m-resolution Swath can contain up to 187,600 data lines. The panchromatic band (band 8) Swath is even larger with a cross-track dimension of 13,200 and an along-track dimension of up to 375,200. Due to the fact that the sizes of the panchromatic band data sets can far exceed the 2 Gb limit on files sizes imposed by HDF, they are segmented into up to three separate Swaths and are written to three separate band files. The LPS writes the band 8 data such that the WRS scenes defined for the subinterval are kept intact and not broken across consecutive segments.

From ephemeris data, the LPS determines where the WRS scenes lie within the Swath and computes the geographic coordinates (longitude and latitude) of each scene's center points and its corners. It then stores the corner coordinates in HDF-EOS Geolocation Fields and maps them to the Swath using the Index. It stores the scene center coordinates in a separate Geolocation Field.

Each scan has an associated time, which is represented in a 64-bit, floating point, *seconds of epoch* form as the number of seconds since 00:00h, 01 January 1993, GMT, rounded to seven decimal places. The LPS writes the time value for each scan to a Geolocation Field and maps it to the Swath. Since there is only one time, but a number of data lines associated with each scan, the mappings of times to data lines in a Swath has a one-to-many cardinality (1:8 for the 60m thermal band, 1:16 for the 30m spectral bands, and 1:32 for the 15m panchromatic band). Other information stored in HDF-EOS Geolocation Fields and associated with each scan, although not technically geolocation data, include the scan number, the scan direction (forward or reverse), and an alphanumeric representation of the scan time, which is based on the year and the day of the year. These data are mapped to the individual data lines with the same one-to-many cardinality as the time values.

Other data are associated with the individual data lines in the Swath. These include the number of each data line, the detector numbers in descending order (i.e., 8…1, 16…1, or 32…1, depending on the band), and the scan data line offsets. The scan data line offsets are of special importance because they indicate where the marginal fill values end and the actual image pixels begin. The LPS provides offset values for each detector for both the right-hand and left-hand sides of each scan. The offset values differ for each data line in a scan because of the physical positioning of the detectors on the ETM+ instrument and must be periodically adjusted to correct for the physical "bumper wear" caused by the scanning motion of the instrument's mirror over time. The LPS writes these data to Geolocation Fields as well.

[**] HDF 4 has a 2 Gb limit on file sizes, which prevents writing the image data from all of the bands to a single file.

The HDF-EOS API supports the subsetting of Swath objects by either longitude or latitude or by time, so long as the Geolocation Fields containing those data are properly mapped to the Swath. Since the Swaths in the Band files are mapped to both temporal and spatial geolocation data, they may be subsetted either by time or by the corner longitude and latitude coordinates of the WRS scenes. In order to subset out segments from the scan or data line based ancillary data, one must compute the start and stop scan numbers or data line numbers of the subset region and read in a slice of the data from the appropriate Geolocation Field.

4.1.2. The Internal Calibrator (IC) Data

The IC data are stored in HDF-EOS Swaths similar to those of the Band files and are also associated with ancillary data stored in Geolocation Fields. A complete LPS product contains only two IC files, one for each format. Unlike the Band files, one IC file contains several Swaths of IC data, one for each band in the format. The much smaller size of the IC data makes this possible. As is the case with the Band data, the lengths of the IC Swaths in the along-track dimension depend upon the duration of the subinterval. The IC data files also contain scan based and data line based ancillary data similar to those in the Band files. Included among these are the scan times in seconds of epoch form, which the LPS writes to an HDF-EOS Geolocation Field. The scan times are mapped to the Swath such that they can be used to define a subset range. The IC data are not mapped to longitude and latitude coordinates or to scenes.

4.1.3. The Mirror Scan Correction Data (MSCD)

The LPS writes the MSCD to an HDF-EOS Point structure. The Point is a sophisticated data structure and the HDF-EOS API provides functions similar to those that it provides for Swaths to subset it. However, the LPS uses it in its most basic form, as essentially just a large table to which it writes one record for each scan in the subinterval. Each record contains twenty-eight fields, including a scan number, a scan time in seconds of epoch format, and a scan time in alphanumeric format. The other twenty-five fields make up the actual mirror scan correction data. The number of records in the Point depends upon the duration of the subinterval. A program can use the HDF-EOS API to subset the MSCD only by time, since the MSCD contains no geographic information and no knowledge of WRS scenes.

The format 1 and format 2 MSCD should be identical, except for a format scan offset caused by a timing discrepancy. Due to this offset, the MSCD for one of the two formats will start and end a few scans earlier than its counterpart and the values in the scan number fields of each record will differ by an amount equal to the offset. Of course, if data breaks occur asymmetrically in the telemetry with respect to the two formats, the two LPS strings may define their subintervals differently and cause the MSCD files to otherwise differ.

4.1.4. The Payload Correction Data (PCD)

The PCD contain the ephemeris and attitude information of the satellite and miscellaneous engineering data. The LPS writes them as a table to a native HDF Vdata structure. Unlike the MSCD, the PCD are not subsettable by the HDF-EOS APIs. The LPS extracts the PCD from the telemetry and converts most of them to engineering units before storing them in the Vdata. Each record in the table contains the PCD for one PCD major frame. A PCD major frame does not correspond to an individual ETM+ scan.

4.1.5. The Multi-Band Browse Files

The Browse files contain viewable color images of the scenes. The LPS produces a Browse file for each of the WRS scenes that it has defined in the format 1 subinterval, including the partial scenes. The operator specifies the bands and their color representations, which the LPS will use in generating the Browse images. The LPS first extracts the data for the given scene from each of the three selected bands. It then performs a special radiometric correction on the data.[6] Next, it applies a waveleting function, which reduces the sizes of the image planes by a factor of 64, while preserving the aspect ratio of the original scene. Contrary to popular belief, the Browse images are not subsampled. The LPS then applies a linear contrast stretch to the images (when processing night scenes, it skips this step). Finally, it passes the images to the HDF *DF24addimage* function, which combines them into a single, pixel-interlaced JPEG-compressed 24-bit color image and writes it out to a RIS24 object in a Browse file. Each Browse file contains the image of one WRS scene. Due to the nature of JPEG compression, the sizes of the final Browse files vary; however, a size of around 200K bytes is typical.

4.1.6. The LPS Metadata

Each LPS string generates metadata when it finishes processing a subinterval. The metadata contain important information about the overall subinterval and about each individual scene. The format 1 and format 2 metadatas are similar, except that the format 1 contains a list of the Browse files and an assessment of cloud cover for each scene. The LPS writes its metadata in Object Description Language (ODL)[7] and stores them in simple text files. ODL is a specialized syntax for organizing structured text, which is readable by both humans and programs.

The subinterval-based metadata contain, among other things, the Swath start and end times, the geographic coordinates of the first and last scans in the Swath, the WRS identifiers of the first and last scenes in the subinterval, fields which indicate which bands are present in the image data, a list of all of the files for the given format included in the LPS product set (except for the Browse files), and the number of WRS scenes defined in the subinterval. There is a separate section in the LPS Metadata files for each scene. Those fields in the scene-based metadata, which relate to the subject of this paper, are the scene's sequence number within the subinterval, an indication if it is a full or partial scene (i.e., contains fewer than 375 scans), the name of the associated Browse file (if format 1), the time of the scene center scan, the longitude and latitude coordinates of the scene center and corners, and the initial gain state (i.e., high-gain or low-gain) of the ETM+ data for each band. The format 1 scene-based metadata also contain an assessment of the amount of cloud cover present in each scene. The LPS uses the Automatic Cloud Cover Assessment (ACCA) algorithm developed by Irish[8] to compute these values.

4.2. The Level Zero-R Products

An L0R product consists of an HDF Directory file and a set of associated external files, which contain the actual product data. It contains a subset of the image data for each band selected for inclusion in the product, a subset of the IC data, a subset of the MSCD, scan data line offset values for the image and IC data, the complete PCD, the complete LPS metadata, a Geolocation Index, the contents of the Calibration Parameter File (CPF), and the ECS Product Metadata. The image and IC data are stored in SDS objects. All of the other types of data are stored in Vdatas. Conveniently, the external files, which contain the metadatas and the contents of the CPF can be viewed independently of the Directory file with any text editor or text file dumper utility. All of these data objects are themselves grouped into Vgroup objects.

4.2.1. The Scan Line Offsets

In the L0R product, the scan line offsets for both the image and IC data are corresponding subsets of those in the initial LPS product; however, they are organized differently. They are stored together as fields in a Vdata along with the associated scan times, scan numbers, scan data line numbers, and detector IDs.

4.2.2. The Geolocation Index

The Geolocation Index is a tabular structure stored in a Vdata, which is analogous to the HDF-EOS Index in the LPS product. The Geolocation Index Vdata contains fields for the corner longitude and latitude coordinates for each scene in the subsetted product; the index numbers of the data lines, which mark the beginnings and endings of each scene in each band resolution in the product; and an indication if the scene is full or partial.

4.2.3. The Calibration Parameter File (CPF)

The CPF is the only component of the L0R product, which is not obtained from the initial LPS product. It contains radiometric and geometric parameters used by the Level 1 processing systems.[9] Like the metadatas, its contents are written in ODL and stored in a Vdata. The Landsat 7 Image Assessment System (IAS)[††] periodically generates a new CPF in order to reflect changes in radiometric and geometric parameters.

4.2.4. The ECS Product Metadata

The ECS Product Metadata contain information specific to the subsetted L0R product. Like the LPS Metadata, they are written in ODL and stored in a Vdata. These metadata include, among other things, the corner geographic coordinates of the data subset; the bounding WRS paths and rows; the number of scans contained in the product; the format scan offset; the

[††] The IAS is part of the Landsat 7 Level 1 processing system.

included bands; the gain states of the first scan of each band; and a list of all of the files, which make up the product (i.e., the names of the HDF Directory file and all of the external element files).

5. GENERATING AN L0R PRODUCT FROM AN LPS PRODUCT

In order to generate an L0R product from an LPS product, a data processing system must first subset out the image, IC, and other relevant ancillary data, reorganize them into native HDF data structures, copy the PCD and the LPS Metadatas, and write them all to a native HDF file set. A discussion of *one* way that this may be done using the HDF-EOS and HDF APIs follows.

5.1. Subsetting Out Single WRS Scenes from the LPS Product

One may subset the HDF-EOS Swaths within the LPS Band files either by specifying an interval of time, which falls within the span of the subinterval, or by specifying a region of the Swath bounded by the Index-mapped WRS scene corner longitude and latitude values. Specifying the subset region of an LPS Band Swath by time interval has several advantages over specifying the region by bounding geographic coordinates. An advantage is that one is not limited to subsetting along the WRS scene boundaries for which there are longitude and latitude values in the Index. Since every scan in the Swath is mapped to a time value, one can subset anywhere within the Swath without concern for the WRS grid. Therefore, partial subintervals, which are not defined along scene boundaries, can also be subsetted out. Another advantage of subsetting by time interval is that one maintains parallelism with the IC and MSCD subsetting. The IC and MSCD are not mapped to geographic coordinates and can only be subsetted by time. To enable subsetting by time interval, the HDF-EOS API provides the *SWdeftimeperiod* function, which specifies the subset time interval and the *SWextractperiod* function, which extracts a subset of the data contained within that time interval from the Swath. An application may use these functions for both the Band and the IC Swaths. The HDF-EOS Point API provides the corresponding functions, *PTdeftimeperiod* and *PTextractperiod*, which an application may use to subset the MSCD.

Regrettably, the other data structures in the LPS product, which need to be subsetted out and included in the L0R product, are not mapped to the scan time Geolocation Field or to the Index. They can still be subsetted out, but this operation requires an extra step. One way that an application might do this is to first identify those scans, which bound the subset interval, by calling the *SWreadfield* function from the HDF-EOS API and passing to it the name of the time field and the total number of items in the structure. This operation will extract the entire time field for the subinterval from the selected Band file into an array. The application can then compare each value in the array to the start and end times of the subset interval. The scans whose times match are the start and end scans. Multiplying the start and end scan numbers (counting from zero) by the data line increment (i.e., the number of data lines per scan) yields the start and end data lines for the subset interval. Once the subsetting application has these values, it can again call the *SWreadfield* function and pass to it the name of the desired field, and the start and end scan or data line numbers. The function will pass back just that slice of the data contained within the specified range. One could use this method to extract both the ancillary scan and data line-based information associated with the subset interval. These ancillary data include the scan times, scan numbers, scan data line numbers, detector Ids, and the scan data line offsets for both the Band and IC data.

5.2. Subsetting Out Multiple WRS Scenes from the LPS Product

Subsetting out multiple scenes adds an extra level of complexity to the process, especially when subsetting the segmented panchromatic band (Band 8) ETM+ data. One must first obtain the start and stop times for each scene in the subset interval from information contained in the band files. The geographic coordinates and band gain states of all WRS scenes in the subinterval can be read from the LPS Metadata files.

In order to subset out multiple WRS scenes from an LPS subinterval, a subsetting application reads the Index to obtain the starting data line of the first scene and the ending data line of the last scene in the subset interval. Given the number of data lines per scan for the given band, it computes the start and end scan numbers for the subset interval. It uses these to obtain the associated times for those scans. These times bound the subset interval. The application then subsets out the data from bands 1-7 in the same manner described for single scene subsetting.

In the case of the panchromatic band (band 8) data, the application first determines if the subset interval is contained completely within one of the segment Swaths. If this is the case, it examines each of the segments to determine which one contains the sought after scenes and subsets it by time. If the application discovers that the subset interval spans across two segments, it determines which two contain the complete time interval. It then subsets from the time of the first scan in the interval in the beginning segment to the last time in that segment. There is overlap between segments, so it computes an average scan period, nominally about 71 milliseconds, and adds it to the time of the last scan in the beginning segment and uses this value as the beginning time value of the interval to be subsetted out of the next segment. It then subsets the next segment from this point to the time of the last scan in the subset interval. Finally, it combines the two data subsets into a single, contiguous SDS and it writes it to the L0R product.

In order to subset the ancillary data, the application reads the time structures from both segments and finds the starting and ending scan numbers of the subset intervals by matching the bounding interval times with scan times. It then extracts the data starting from the scan in the first segment whose time matches the interval start time to the last scan in that segment. It adds an average scan period to the last scan's time and subsets the second segment from that scan to the scan whose time matches the end of the subset interval. It combines these data into a single structure and writes them to a Vdata in the L0R product. The LPS does not segment band 8 IC data or MSCD; therefore, the application subsets those data in the same manner as it does the other bands.

As in the case of single scene subsetting, the application can obtain the geographic coordinates of the corners of each scene from the LPS Metadata files. It writes these to the Geolocation Index Vdata in the L0R product. The geographic coordinates of the full L0R product's corners are, of course, just the upper-left and upper-right corners of the first scene in the subset and the lower-left and lower right corners of the last scene in the subset. It writes these values to the ECS Product Metadata.

5.3. Subsetting Out Partial Subintervals from the LPS Product

Partial subinterval products are similar to multiple WRS scene products. Except for the fact that the first and last scenes in a partial subinterval product will usually be partial scenes as indicated in the Geolocation Index, an end user will notice no differences between the two product types.

There is one area where partial subinterval subsetting requires some special processing. The ECS Product Metadata contains the corner longitude and latitude coordinates of the subsetted product. In the case of WRS scene based products, the corner coordinates of the first scan of the first scene and the last scan of the last scene are the corner coordinates of the full product. The LPS supplies these data. However, the LPS only provides geographic coordinates for the WRS scenes it has defined. Geographic coordinates are not available on an individual scan basis. Since the start and end scans in partial subintervals do not lie on WRS scene boundaries or centers, bounding geographic coordinates are not available for the overall products. In order to provide values to the necessary fields in the ECS Product Metadata and to the corner coordinate fields of the first and last scene records in the Geolocation Index, the subsetting application must interpolate the product geographic coordinates. Because the Earth is spheroidal and because the path of the Landsat 7 satellite does not exactly follow the meridians of longitude, a simple linear interpolation of the product geographic coordinates from the available WRS scene coordinates is not possible. A more sophisticated algorithm must be used.

6. AN EXAMPLE OF A SUBSETTING APPLICATION

The DAAC Emergency System (DES) and the SubsetLPS tool are two applications that implement the subsetting and product formatting method described in this paper. The DES is a single-scene L0R product generation system that a team at Computer Sciences Corporation and Goddard Space Flight Center (GSFC) developed in 1998 for the EDC. The SubsetLPS tool is a stand-alone application, which the author developed in 1999 in response to the need for an easy way to produce single-scene, multiple-scene, and partial subinterval-based L0R products from LPS output in a testing and analysis environment. Much of the code in the SubsetLPS tool was reused from the DES.

To generate partial subinterval products, the SubsetLPS tool allows the user to define the desired subset range within a subinterval by specifying either a time interval (in alphanumeric form), start and end scan numbers, bounding longitudes and latitudes, or fractions of WRS scenes. In order to obtain product corner coordinates for partial subintervals, the SubsetLPS tool implements an interpolation algorithm developed by Boia,[10] which computes the longitude and latitude coordinates of any selected scan. The Boia algorithm uses the known coordinates and scan numbers of the center scans of two selected WRS scenes as references. Through these reference points, it defines great circle arcs along the surface of the Earth. It plots the midpoints of the reference scans and the scan of interest along the arcs, and interpolates the geographic coordinates of that scan by computing angular distances between those points. This algorithm can also be applied in reverse to find the scan within a Swath, which is nearest to a set of longitude and latitude coordinates. The SubsetLPS tool implements this in order to allow the user to define subset intervals by specifying beginning and ending geographic coordinates.

7. SUMMARY AND CONCLUSIONS

The Landsat 7 science data consist of multi-band spectral, thermal, and panchromatic images. The LPS extracts these images from encoded telemetry, applies scan alignment corrections, geolocates them, and assembles them into subinterval-based, Level 0 data products. It then stores them in HDF-EOS Swath and Point data objects along with ancillary information. The Band data, IC data, and MSCD must first be subsetted and reformatted into native HDF-based L0R products prior to their ingestion by the Level 1 processing systems. There is a one-to-one mapping of time values to ETM+ scans in the LPS products. An application can create various configurations of L0R products by defining the desired Level 0R data sets in terms of time intervals and calling the HDF-EOS API to subset them out by time period. Subsetting by time provides for a much finer resolution than defining the subset interval by the geographic coordinates of WRS scene corners. It also simplifies the extraction of partial subintervals for which there are no associated geographic data.

In order to take full advantage of HDF-EOS, the author recommends that future Landsat-type data products, if they are to be stored in HDF-EOS format, should write all of their ancillary data, not just the instrument data, to Swaths or Points, which are mapped to time-based and geographic coordinate-based Geolocation Fields. This would allow a product to be subsetted entirely through the HDF-EOS subsetting functions and would eliminate the intermediate steps of having to read the whole time field into an array in order to match the values with scan numbers. The ability to map time values and geographic coordinates directly to image and point data make HDF-EOS a convenient format for storing Earth science data collected by orbiting satellites, such as Landsat 7. HDF's platform independence and ability to manage and section large, multidimensional arrays make it a good choice for storing science data in general.

ACKNOWLEDGEMENTS

The SubsetLPS tool was developed for the National Aeronautics and Space Administration (NASA), Goddard Space Flight Center (GSFC), Greenbelt, Maryland as part of the Landsat 7 project under contract NAS5-31500. The author wishes to thank Robert Schweiss of NASA and Nathaniel Daniel, Neil Ottenstein, Deborah Derrick, and Janice Swope of Computer Sciences Corporation for their support and for reviewing a draft version of this manuscript.

REFERENCES

1. National Aeronautics and Space Administration (NASA), Goddard Space Flight Center (GSFC), *Landsat 7 System Zero-R Distribution Product Data Format Control Book*, Revision 3, 430-11-06-007-3, May 1999.
2. National Aeronautics and Space Administration (NASA), Goddard Space Flight Center (GSFC), *Landsat 7 Processing System (LPS) Output Files Data Format Control Book*, Revision 5, 510-3FCD/0195, July 1999.
3. Univ. of Illinois at Urbana-Champaign, National Center for Supercomputing Applications (NCSA), *HDF User's Guide*, Ver. 4.1r1, May 1997.
4. Hughes Information Technology Systems, *HDF-EOS Library User's Guide for the ECS Project, Volume 1: Overview and Examples*, 170-TP-005-003, Upper Marlboro, Maryland, April 1997.
5. B. Fortner, "HDF: the hierarchical data format," *Dr. Dobb's Journal*, pp. 42-48, May 1998.
6. B. Markam and J. Barker, "Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures," *EOSAT Technical Notes*, 1, pp. 3-8, 1986.
7. Consultive Committee on Space Data Systems (CCSDS), *Report Concerning Space Data System Standards: Parameter Value Language - A Tutorial*, Green Book, Issue 1, CCSDS 641.0-G-1, May 1992.
8. R. Irish, "Landsat 7 automatic cloud cover assessment," in *Proceedings of SPIE*, Vol. 4049, (in press), 2000.
9. National Aeronautics and Space Administration (NASA), Goddard Space Flight Center (GSFC), *Landsat 7 Calibration Parameter File Definition*, Revision 3, 430-15-01-002-3, April 1999.
10. J. Boia, "An algorithm for estimating ETM+ scan number as a function of longitude and latitude," *unpublished document*, May 1999.